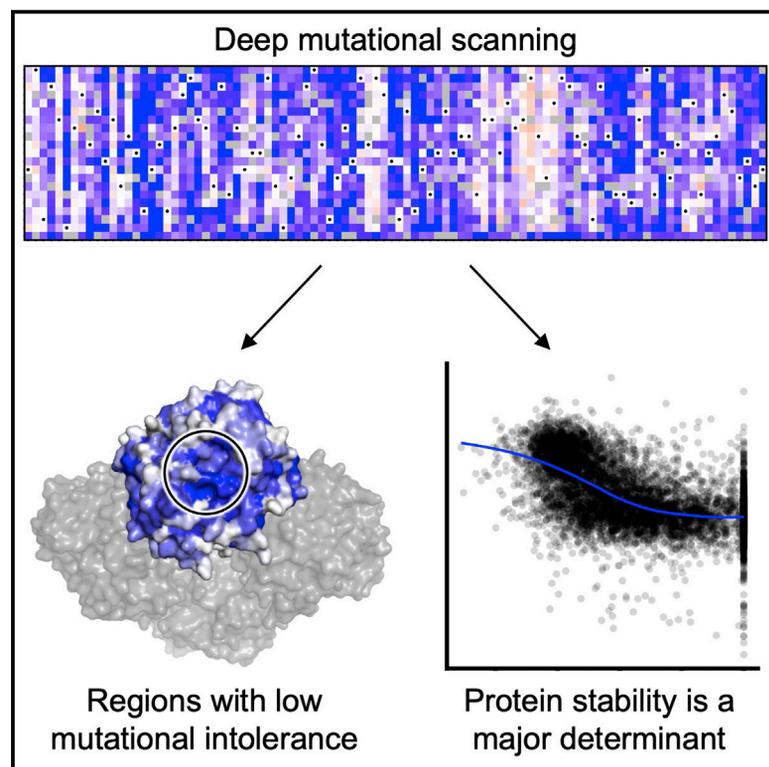


Mutational fitness landscape of human influenza H3N2 neuraminidase

Graphical abstract



Authors

Ruipeng Lei, Andrea Hernandez Garcia, Timothy J.C. Tan, ..., Satish K. Nair, Jian Peng, Nicholas C. Wu

Correspondence

jianpeng@illinois.edu (J.P.),
nicwu@illinois.edu (N.C.W.)

In brief

Lei et al. measure the replication fitness effects of >6,000 single amino acid mutations in human influenza H3N2 neuraminidase, which is emerging as a vaccine target for broadened protection. The results provide insights into the evolutionary potential and constraints of influenza neuraminidase.

Highlights

- Deep mutational scanning (DMS) of human influenza H3N2 neuraminidase head domain
- DMS result correlates with mutational fitness inferred from natural sequences
- Mutational fitness effects correlate with protein stability effects ($\Delta\Delta G$)



Report

Mutational fitness landscape of human influenza H3N2 neuraminidase

Ruipeng Lei,¹ Andrea Hernandez Garcia,¹ Timothy J.C. Tan,² Qi Wen Teo,^{1,3} Yiquan Wang,¹ Xiwen Zhang,⁴ Shitong Luo,⁴ Satish K. Nair,^{1,2,3} Jian Peng,^{4,5,*} and Nicholas C. Wu^{1,2,3,6,7,*}

¹Department of Biochemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

²Center for Biophysics and Quantitative Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

³Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

⁴Helixon Limited, Beijing 100084, China

⁵Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

⁶Carle Illinois College of Medicine, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

⁷Lead contact

*Correspondence: jianpeng@illinois.edu (J.P.), nicwu@illinois.edu (N.C.W.)

<https://doi.org/10.1016/j.celrep.2022.111951>

SUMMARY

Influenza neuraminidase (NA) has received increasing attention as an effective vaccine target. However, its mutational tolerance is not well characterized. Here, the fitness effects of >6,000 mutations in human H3N2 NA are probed using deep mutational scanning. Our result shows that while its antigenic regions have high mutational tolerance, there are solvent-exposed regions with low mutational tolerance. We also find that protein stability is a major determinant of NA mutational fitness. The deep mutational scanning result correlates well with mutational fitness inferred from natural sequences using a protein language model, substantiating the relevance of our findings to the natural evolution of circulating strains. Additional analysis further suggests that human H3N2 NA is far from running out of mutations despite already evolving for >50 years. Overall, this study advances our understanding of the evolutionary potential of NA and the underlying biophysical constraints, which in turn provide insights into NA-based vaccine design.

INTRODUCTION

Despite the practice of social distancing and masking during the COVID-19 pandemic, influenza A virus continues to circulate in the human population. In fact, influenza H3N2 virus has been circulating in human for more than half a century without any sign of disappearing. Influenza A virus has two surface glycoprotein antigens, hemagglutinin (HA) and neuraminidase (NA). Compared with HA, which has traditionally been the focus of influenza vaccine development, NA is poorly characterized. Nevertheless, NA is emerging as an effective vaccine target due to studies that demonstrate anti-NA immunity as an independent correlate of protection^{1–4} as well as the recent discovery of broadly protective NA antibodies.⁵

NA is a homotetrameric type II transmembrane protein with an N-terminal cytoplasmic tail, a transmembrane domain, a hyper-variable stalk region, and a C-terminal head domain. The head domain, which accounts for ~80% of NA protein sequence, contains an enzymatic active site that cleaves sialic acid to facilitate virus release and prevent virion aggregation.⁶ The enzymatic active site of influenza NA is also a target of several FDA-approved antivirals.⁷ Similar to HA, NA of human influenza A virus also undergoes antigenic drift.^{8,9} The major antigenic regions of NA are at the rim of its enzymatic active site.^{10,11} Our previous study on seven antigenic residues has demonstrated that charge balancing constrains the antigenic evolution of NA.¹² More

recently, we have also shown that epistasis between certain natural mutations of human H3N2 NA can be explained by protein stability.¹³ While these studies have helped understand the biophysical constraints of NA evolution, the mutational fitness landscape of NA remains largely elusive.

The development of deep mutational scanning, which combines saturation mutagenesis and next-generation sequencing, enables many mutations to be studied in a massively parallel manner and thus allows the mutational fitness landscape to be empirically determined.¹⁴ Recent deep mutational scanning studies of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) receptor-binding domain have provided critical insights into the evolution and antibody escape of SARS-CoV-2^{15–17} as well as COVID-19 vaccine design.¹⁸ Similarly, deep mutational scanning has also been successfully applied to probe the mutational fitness landscape and antibody escape of influenza HA.^{19–24} However, most, if not all, deep mutational scanning studies of viral proteins were performed *in vitro* (i.e., non-animal systems). The correlation between *in vitro* fitness measurement and *in vivo* replication or transmission fitness is not well characterized but is fundamental to virus research.

In this study, we used deep mutational scanning to measure the fitness effects of >6,000 single amino acid mutations in the head domain of influenza H3N2 NA. The result enabled us to identify solvent-exposed regions that have low mutational tolerance, which are critical to vaccine development. Our analysis



also demonstrated that protein stability is a major determinant of the mutational fitness landscape of NA. In addition, fitness measurements in our deep mutational scanning experiment, which was performed *in vitro*, correlated well with mutational fitness that was inferred from the natural sequences of human H3N2 NA (Spearman's rank correlation = 0.59). This observation further allowed us to draw conclusion about the evolutionary potential of NA in circulating human H3N2 virus.

RESULTS

Deep mutational scanning of human H3N2 NA

We aimed to analyze the fitness effects of mutations in the NA of H3N2 A/Moscow/10/1999 (Mos99) using deep mutational scanning (Figure S1). Briefly, saturation mutagenesis was applied to generate all possible single amino acid mutations from residues 82 to 465 in Mos99 NA, which covered 99% of the head domain (residues 82 to 469). Of note, N2 numbering is used unless otherwise stated. The virus mutant library was then rescued by the influenza reverse-genetics system²⁵ and passaged once in MDCK-SIAT1 cells, which could minimize the emergence of cell-adaptive mutation in NA.²⁶ Using next-generation sequencing, the frequencies of individual mutations in the plasmid mutant library and the post-passaged mutant library were measured. Subsequently, the fitness of each mutation was computed as the normalized \log_{10} enrichment ratio such that the mean fitness value of silent mutations was 0, whereas beneficial and deleterious mutations would have positive and negative fitness values, respectively.

Our deep mutational scanning measured the fitness effects of 6,353 (87%) out of 7,296 all possible amino acid mutations across the 384 residues of interest (Figure 1; Table S1). A Pearson correlation of 0.84 was obtained between two biological replicates (Figure S2A), which was on the high end among deep mutational experiments using the recombinant influenza virus.^{20,21,27,28} The deep mutational scanning result here also correlated well with our previous study that focused on a seven-residue antigenic region in Mos99 NA¹² (Figure S2B). Besides, the fitness distributions of silent and nonsense mutations had minimal overlap, further validating that fitness selection has taken place (Figure S2C).

Biophysical determinants of mutational tolerance

We aimed to identify the biophysical constraints that determine the mutational tolerance of the NA head domain. Here, we defined the mutational tolerance at each residue as the average fitness of mutations at the given residue. Residues in the active site and antigenic regions were defined as previously described.^{6,10} Using the crystal structure of Mos99 NA (PDB: 7U4F) that was determined in our recent study,¹³ the remaining residues were categorized according to relative solvent accessibility (RSA). Residues in the protein core, the protomer-protomer interface, and surface were categorized as “buried,” “interface,” and “exposed,” respectively (see STAR Methods).

Mutational tolerance of the exposed residues and antigenic regions was significantly higher than that of buried residues, interface residues, and active site residues ($p \leq 0.0001$) (Figures 2A–2C). In contrast, mutational tolerance of the buried residues,

interface residues, and active site residues did not differ significantly ($p \geq 0.19$). Similarly, mutational tolerance of the exposed residues and antigenic regions did not differ significantly ($p = 0.73$). The mutational tolerance of different residue categories appeared to associate with their RSA. For example, residue categories with higher mutational tolerance, namely the exposed residues and antigenic regions, have higher RSA (Figure S3A). In fact, mutational tolerance correlated well with RSA for the 384 residues that were examined in our deep mutational scanning experiment (Spearman's rank correlation = 0.66) (Figure 2D), consistent with similar studies on other proteins.^{15,19} Since mutations at residues with lower RSA are typically more destabilizing,^{29–31} our results indicate that folding and tetramerization stabilities are the major determinants of mutational tolerance in the NA head domain.

To further examine if the active site constrained the mutational tolerance of the surrounding residues, we determined the crystal structure of Mos99 NA in complex with sialic acid at 1.6 Å resolution (Table S2). Although the sialidase enzymatic activity is the major function of NA, mutational tolerance and the distance from active site only showed weak correlation (Spearman's rank correlation = 0.18) (Figure S3B). The correlation did not improve if we only included those residues within 15 Å from the active site (Spearman's rank correlation = 0.03) (Figure S3D). These observations substantiate the notion that optimal enzymatic activity of NA relies on the global conformational stability of the tetramer.^{6,32,33}

Clusters of surface residues with low mutational tolerance

When we projected the mutational tolerance data on the structure, two clusters of solvent-exposed residues with low mutational tolerance were identified (Figures 2C and 3A). The first cluster was the active site (see above; Figure 2C), which was expected due to its functional importance. The second cluster was at the membrane proximal side of the NA head domain (Figure 3A), consisting of five residues with low mutational tolerance, namely R283, R288, D304, D355, and W383 (Figure 3B). Together, these five residues have a solvent exposure surface area of 222 Å² and form many interactions among themselves. Specifically, R283, R288, D304, and D355 form a network of electrostatic interactions. Furthermore, R288 forms π -stacking interactions with R283 and W383. As a result, the low mutational tolerance of these five residues is likely due to their critical role in NA folding stability. Consistently, these five residues are highly conserved in human H3N2 NA, albeit not across NA subtypes (Figure 3C).

Correlation between protein stability and mutational fitness effects

To systematically investigate how protein stability of NA influences the mutational fitness effects, we used FoldX^{34,35} to predict the protein stability effects of the 6,353 amino acid mutations that were examined in our deep mutational scanning experiment (Table S3). We observed a non-linear relationship between predicted mutational stability ($\Delta\Delta G$) of the NA tetramer and mutational fitness (Figure 4A). Mutations that were predicted to be more stabilizing ($\Delta\Delta G < 0$) tended to have a higher fitness. As the predicted stability decreased (i.e., $\Delta\Delta G$ increased), mutational fitness decreased and reached a plateau at a fitness of around -1 ,

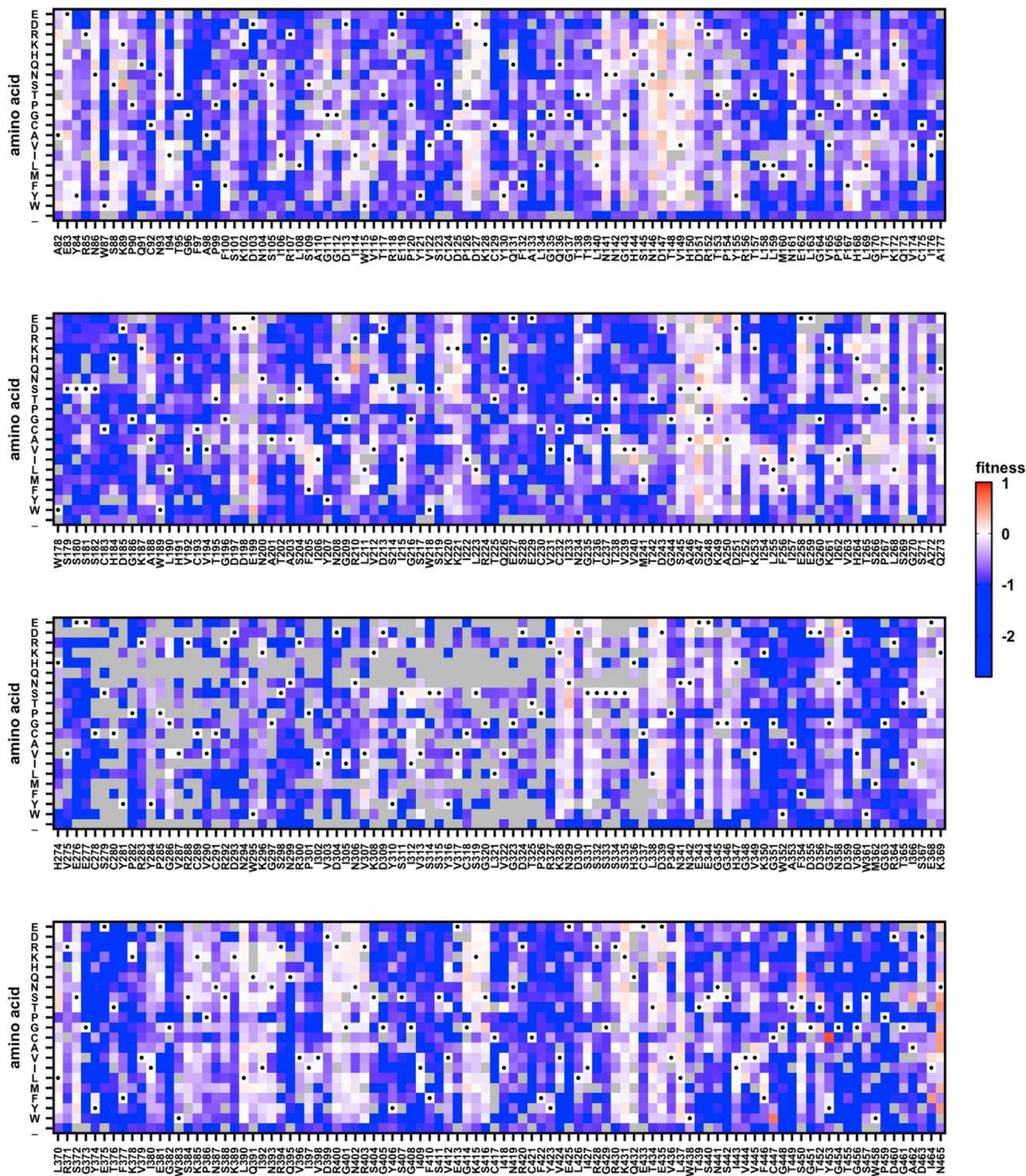


Figure 1. Deep mutational scanning of Mos99 NA head domain

The fitness of individual mutations in the Mos99 NA head domain was measured by deep mutational scanning and is shown as a heatmap. Wild-type (WT) amino acids are indicated by a black circle. Mutations in gray were excluded in our data analysis due to low input count or high occurrence in the WT sample (see [STAR Methods](#)).

See also [Figures S1](#) and [S2](#) and [Table S1](#).

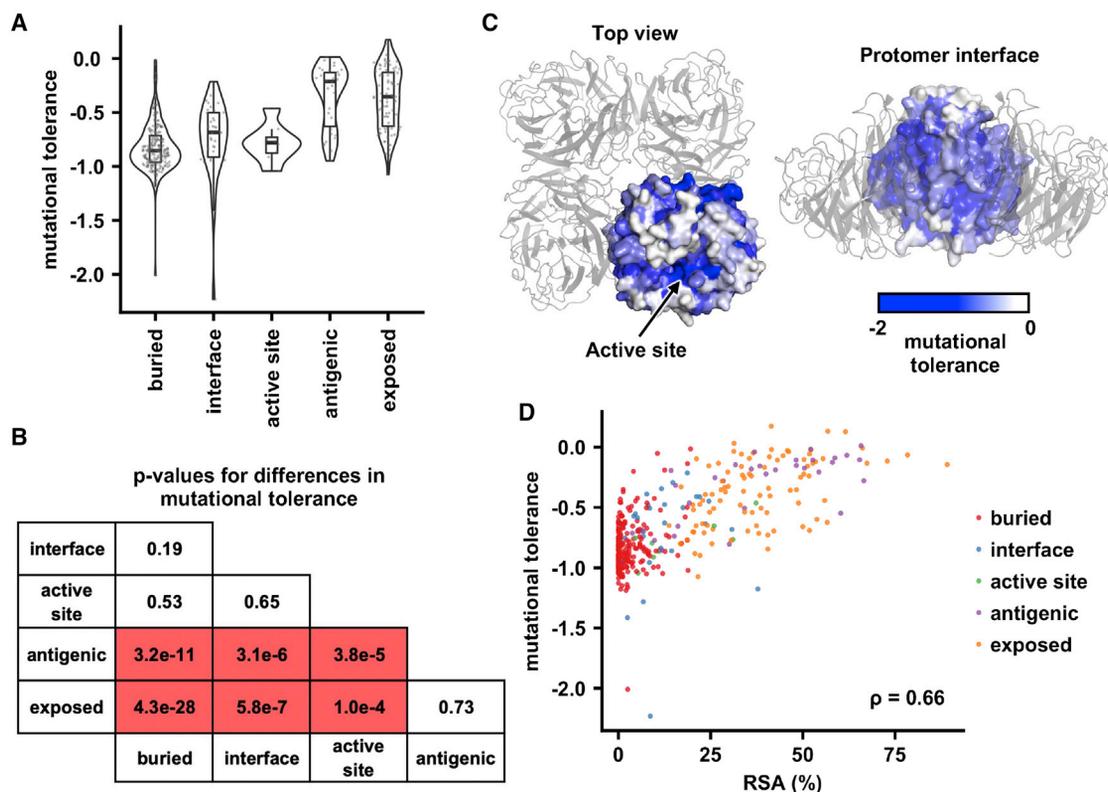


Figure 2. Biophysical determinants of mutational tolerance

(A) Mutational tolerance was computed by averaging the fitness of mutations at each residue. The distribution of mutational tolerance for each residue category is shown as a violin plot and a boxplot. For the boxplot, the middle horizontal line represents the median. The lower and upper hinges represent the first and third quartiles, respectively. The upper whisker extends to the highest data point within a $1.5 \times$ inter-quartile range (IQR) of the third quartile, whereas the lower whisker extends to the lowest data point within a $1.5 \times$ IQR of the first quartile. Each data point represents the mutational tolerance of one residue.

(B) The differences in mutational tolerance between residue categories were computed by two-tailed Student's *t* tests. The *p* value for each pairwise comparison is shown.

(C) Mutational tolerance of each residue is shown on Mos99 NA (PDB: 7U4F).¹³ One protomer is in surface representation with individual residues colored according to their mutational tolerance. The other three protomers are in gray cartoon representation.

(D) The relationship between relative solvent accessibility (RSA) of each residue on the Mos99 NA tetramer (PDB: 7U4F)¹³ and its mutational tolerance is shown. Each data point represents one residue and colored according to the residue category. The Spearman's rank correlation coefficient (ρ) is indicated.

See also Figure S3 and Table S2.

which was similar to the average fitness of nonsense mutations (Figure S2C). This observation indicates that once the NA protein is sufficiently destabilized, the virus would not be able to replicate. The correlation between predicted $\Delta\Delta G$ and mutational fitness (Spearman's rank correlation = -0.61) confirms that protein stability is a major determinant of NA mutational fitness effects. Similar observations were made when a single NA protomer (i.e., monomeric form) was used for computing the predicted mutational stability, despite a slightly lower correlation with mutational fitness (Spearman's rank correlation = -0.58) (Figure S3C). Overall, these results substantiate the notion that protein stability can help predict the functional effects of mutations³⁶ and represents a major constraint of molecular evolution.^{37–39}

Correlation between natural evolution and mutational fitness effects

Next, we would like to examine if our deep mutational scanning result reflects the mutational fitness in the natural evolution of

human H3N2 NA. Recently, the development of MSA transformer, which is a deep-learning-based protein language model, has allowed unsupervised inference of mutational fitness for any protein using multiple sequence alignment.^{40–42} Here, we applied MSA transformer to infer the fitness of mutations in the Mos99 NA using 66,562 NA sequences from human H3N2 strains that were isolated between 1968 and 2020. The inferred mutational fitness would therefore represent the mutational fitness in the natural evolution of human H3N2 NA (Table S3). For the 6,353 mutations that were examined in our deep mutational scanning experiment, the inferred fitness correlated well with the fitness measurement by deep mutational scanning (Spearman's rank correlation = 0.59) (Figure 4B). The correlation was higher if only naturally observed mutations were considered ($n = 1,391$, Spearman's rank correlation = 0.69) (Figure 4C). A moderate correlation could still be obtained if we only considered those mutations that were not observed in nature ($n = 4,962$, Spearman's rank correlation = 0.48) (Figure 4D). A

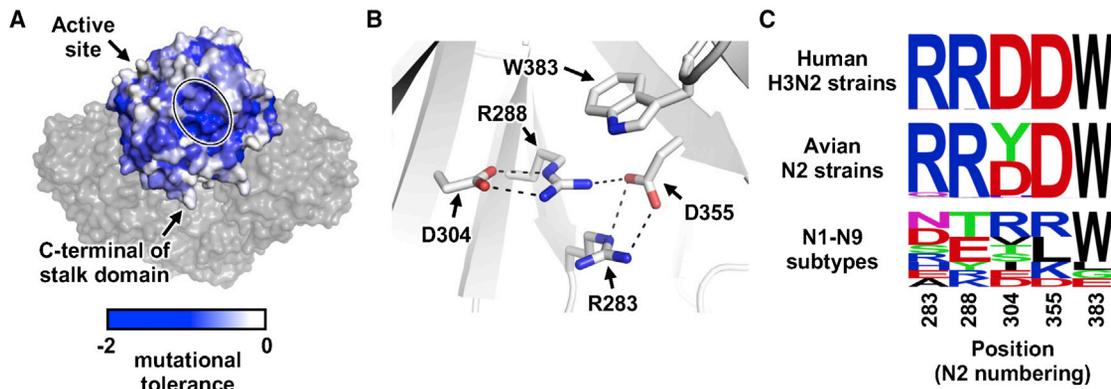


Figure 3. A cluster of residues with low mutational tolerance at the membrane proximal side of the NA head domain

(A) Mutational tolerance of each residue is shown on Mos99 NA (PDB: 7U4F).¹³ One protomer is in surface representation with individual residues colored according to their mutational tolerance. The other three protomers are in gray surface representation. The C-terminal of the stalk domain, which connects the NA head domain to the membrane, is indicated. A cluster of residues with low mutational tolerance is circled.

(B) Residues within the cluster in (A) are shown. Electrostatic interactions are represented by dashed lines.

(C) The natural variants of NA residues 283, 288, 304, 355, and 383 (N2 numbering) are represented by sequence logos. Top: 58,937 human H3N2 strains from 1968 to 2020. Middle: 6,900 avian N2 strains. Bottom: representative strains from N1-N9 subtypes.

See also Table S2.

handful of naturally observed mutations had a high inferred fitness but were neutral in our deep mutation scanning experiment (horizontal tail in the top right of Figure 4C), potentially due to the positive selection pressures that were absent *in vitro*, such as herd immunity and transmission efficiency. Overall, despite the differences in selection pressures that are seen *in vitro* and in human population, our deep mutational scanning result largely aligns with the mutational fitness of human H3N2 NA in the natural setting.

Unexplored functional sequence space in human H3N2 NA

As mentioned above, among the 6,353 mutations that were examined in our deep mutational scanning experiment, only 1,390 (22%) have been observed in naturally circulating human H3N2 NA. We aimed to assess if any of the remaining 4,963 (78%) mutations have the potential to emerge in circulating human H3N2 virus. According to our deep mutational scanning result, while the majority of these 4,963 mutations were deleterious to the virus, 251 had a fitness >0 (Figure 4E). Of note, mutations with fitness <0 have also been observed in nature (Figures 4C and 4E), although most of the highly deleterious ones (fitness < -1) are with low occurrence frequency (Figure S4). Given that our deep mutational scanning result correlates well with the mutational fitness in the natural setting (see above; Figure 4B), these observations suggest that circulating human H3N2 NA is far from running out of mutations.

DISCUSSION

While NA is an emerging target for influenza vaccine development, the biophysical constraints on its evolution have not been entirely clear. In this study, we applied deep mutational scanning to study the mutational fitness landscape of influenza H3N2 NA. One important finding is that the *in vitro* fitness mea-

surement in our deep mutational scanning experiment correlates well with the *in vivo* mutational fitness that is inferred from natural sequences. This observation has important implications for how to interpret *in vitro* fitness data of influenza NA mutants. However, we acknowledge that future studies need to examine if this correlation can be observed in other viral proteins.

HA and NA in human influenza virus undergoes antigenic drift independently.^{8,9} Previous deep mutational scanning studies have shown that the antigenic regions of HA have high mutational tolerance.^{19,21} Similarly, our work here demonstrates that the antigenic regions of NA also have high mutational tolerance. This result indicates that a broadly protective NA-based vaccine should focus the antibody response against immunosubdominant regions that have lower mutational tolerance. In fact, two clusters of solvent-accessible residues with low mutational tolerance are observed in this study. The first cluster is the active site, which is known to be highly conserved across influenza A and B⁶ and is a target of broadly protective antibodies.⁵ The second cluster, which is at the membrane-proximal side of the NA head domain, is highly conserved within the N2 subtype but not across different subtypes. Antibodies to this region, if any, should be at least somewhat cross-protective against antigenically distinct N2 strains and have a higher genetic barrier to escape compared with those that target the classic antigenic regions of NA. As the antigenicity of NA remains poorly understood, future studies should explore whether this second cluster of residues are targeted by human antibodies.

Recently, Moderna has initiated a phase 1/2 clinical trial of influenza mRNA vaccines that contain NA as a component (ClinicalTrials.gov: NCT05333289). The increasing importance of NA in public health urges continued efforts to characterize the evolutionary biology and biophysical properties of influenza NA, which will in turn facilitate the optimization of NA-based immunogens. The conformational stability of an immunogen often represents a key to achieve high vaccine

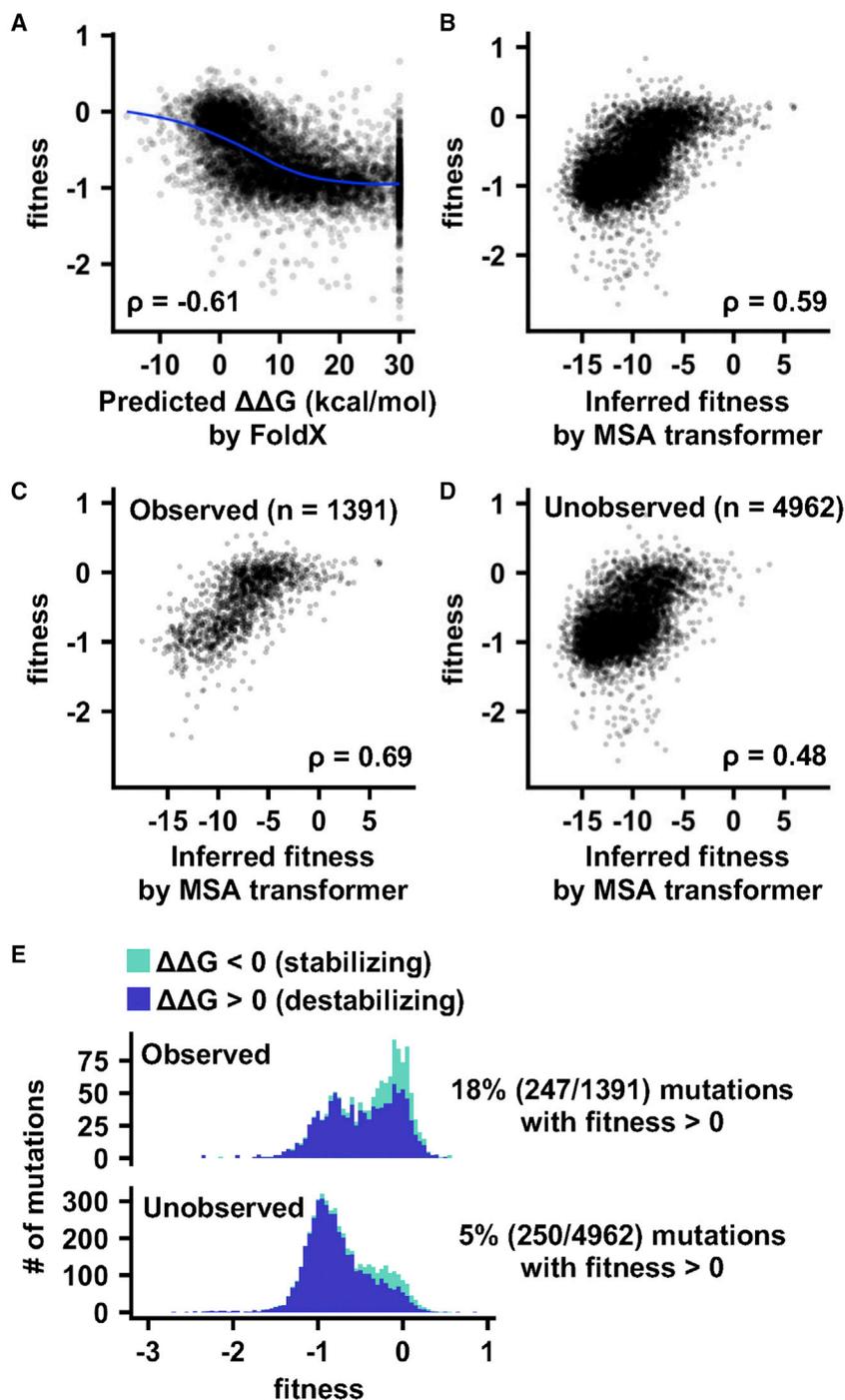


Figure 4. Deep mutational scanning result correlates with mutational stability and natural evolution

(A) The relationship between predicted stability effect ($\Delta\Delta G$) by FoldX^{34,35} and fitness measurement by deep mutational scanning is shown. Mutations with predicted $\Delta\Delta G > 30$ kcal mol⁻¹ are shown as 30 kcal mol⁻¹. A smooth curve was fitted by loess. (B) The relationship between fitness inferred from natural sequences using MSA transformer⁴⁰ and fitness measured by deep mutational scanning is shown. The Spearman's rank correlation coefficient (ρ) is indicated.

(C and D) Same as (B) except only those mutations that are (C) observed in nature or (D) not observed in nature are shown.

(E) Distributions of fitness effects of mutations that have been observed in nature (observed) and those that have not been observed in nature (unobserved) are shown.

See also Figure S4 and Tables S2 and S3.

effectiveness.⁴³ Using deep mutational scanning data to identify stabilizing mutations has been demonstrated in a study on the SARS-CoV-2 receptor-binding domain.¹⁸ Since studies have shown that the conformational stability of NA can be improved by stabilizing mutations,^{44,45} we anticipate that our deep mutational scanning dataset will be instrumental for NA-based immunogen design.

Limitations of the study

In this study, the Mos99 NA mutant library was rescued using the six internal segments from A/WSN/33 (H1N1) and HA from A/Hong Kong/1/1968 (H3N2). While the internal segments are unlikely to influence the mutational fitness landscape of NA, it is known that HA and NA functionally interact.^{46–48} Consistently, our recent study has shown that the mutational fitness

landscape of HA can be influenced by NA.⁴⁹ As a result, we acknowledge that our deep mutational scanning experiment may be slightly different if both HA and NA from Mos99 are used.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
 - Lead contact
 - Materials availability
 - Data and code availability
- **EXPERIMENTAL MODELS AND SUBJECT DETAILS**
 - Cell cultures
- **METHOD DETAILS**
 - Mutant library construction
 - Rescuing and passaging the mutant library
 - Sequencing library preparation
 - Sequencing data analysis
 - Protein expression and purification
 - Crystallization and structural determination
 - Residue classification
 - Sequence logo analysis
 - FoldX
 - MSA transformer
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.celrep.2022.111951>.

ACKNOWLEDGMENTS

This work was supported by the Searle Scholars Program (N.C.W.) and National Institutes of Health (NIH) R01 AI165475 (N.C.W.).

AUTHOR CONTRIBUTIONS

R.L., J.P., and N.C.W. conceived and designed the study. R.L. constructed the mutant libraries and prepared the sequencing libraries. T.J.C.T. and Q.W.T. assisted with experiments. Y.W. and N.C.W. analyzed the next-generation sequencing data. R.L., A.H.G., and S.K.N. performed protein crystallization. A.H.G. and S.K.N. collected the X-ray diffraction data as well as determined and refined the structures. X.Z., S.L., and J.P. performed FoldX analysis and fitness inference using MSA transformer. R.L. and N.C.W. wrote the paper and all authors reviewed and edited the paper.

DECLARATION OF INTERESTS

X.Z., S.L., and J.P. are employees of HeliXon. N.C.W. consults for HeliXon.

Received: August 19, 2022

Revised: November 24, 2022

Accepted: December 16, 2022

REFERENCES

1. Monto, A.S., Petrie, J.G., Cross, R.T., Johnson, E., Liu, M., Zhong, W., Levine, M., Katz, J.M., and Ohmit, S.E. (2015). Antibody to influenza virus neuraminidase: an independent correlate of protection. *J. Infect. Dis.* 212, 1191–1199. <https://doi.org/10.1093/infdis/jiv195>.
2. Couch, R.B., Atmar, R.L., Franco, L.M., Quarles, J.M., Wells, J., Arden, N., Niño, D., and Belmont, J.W. (2013). Antibody correlates and predictors of immunity to naturally occurring influenza in humans and the importance of antibody to the neuraminidase. *J. Infect. Dis.* 207, 974–981. <https://doi.org/10.1093/infdis/jis935>.
3. Memoli, M.J., Shaw, P.A., Han, A., Czajkowski, L., Reed, S., Athota, R., Bristol, T., Fargis, S., Risos, K., Powers, J.H., et al. (2016). Evaluation of anti-hemagglutinin and antineuraminidase antibodies as correlates of protection in an influenza A/H1N1 virus healthy human challenge model. *mBio* 7, e00417-16. <https://doi.org/10.1128/mBio.00417-16>.
4. Weiss, C.D., Wang, W., Lu, Y., Billings, M., Eick-Cost, A., Couzens, L., Sanchez, J.L., Hawksworth, A.W., Seguin, P., Myers, C.A., et al. (2020). Neutralizing and neuraminidase antibodies correlate with protection against influenza during a late season A/H3N2 outbreak among unvaccinated military recruits. *Clin. Infect. Dis.* 71, 3096–3102. <https://doi.org/10.1093/cid/ciz1198>.
5. Stadlbauer, D., Zhu, X., McMahon, M., Turner, J.S., Wohlbold, T.J., Schmitz, A.J., Strohmeier, S., Yu, W., Nachbagauer, R., Mudd, P.A., et al. (2019). Broadly protective human antibodies that target the active site of influenza virus neuraminidase. *Science* 366, 499–504. <https://doi.org/10.1126/science.aay0678>.
6. McAuley, J.L., Gilbertson, B.P., Trifkovic, S., Brown, L.E., and McKimm-Breschkin, J.L. (2019). Influenza virus neuraminidase structure and functions. *Front. Microbiol.* 10, 39. <https://doi.org/10.3389/fmicb.2019.00039>.
7. CDC (2022). Antiviral drugs for seasonal influenza: additional links and resources. <https://www.cdc.gov/flu/professionals/antivirals/links.htm>.
8. Kilbourne, E.D., Johansson, B.E., and Grajower, B. (1990). Independent and disparate evolution in nature of influenza A virus hemagglutinin and neuraminidase glycoproteins. *Proc. Natl. Acad. Sci. USA* 87, 786–790.
9. Sandbulte, M.R., Westgeest, K.B., Gao, J., Xu, X., Klimov, A.I., Russell, C.A., Burke, D.F., Smith, D.J., Fouchier, R.A.M., and Eichelberger, M.C. (2011). Discordant antigenic drift of neuraminidase and hemagglutinin in H1N1 and H3N2 influenza viruses. *Proc. Natl. Acad. Sci. USA* 108, 20748–20753. <https://doi.org/10.1073/pnas.1113801108>.
10. Colman, P.M., Varghese, J.N., and Laver, W.G. (1983). Structure of the catalytic and antigenic sites in influenza virus neuraminidase. *Nature* 303, 41–44. <https://doi.org/10.1038/303041a0>.
11. Air, G.M. (2012). Influenza neuraminidase. *Influenza Other Respir. Viruses* 6, 245–256. <https://doi.org/10.1111/j.1750-2659.2011.00304.x>.
12. Wang, Y., Lei, R., Nourmohammad, A., and Wu, N.C. (2021). Antigenic evolution of human influenza H3N2 neuraminidase is constrained by charge balancing. *Elife* 10, e72516. <https://doi.org/10.7554/eLife.72516>.
13. Lei, R., Tan, T.J.C., Hernandez Garcia, A., Wang, Y., Diefenbacher, M., Teo, C., Gopan, G., Tavakoli Dargani, Z., Teo, Q.W., Graham, C.S., et al. (2022). Prevalence and mechanisms of evolutionary contingency in human influenza H3N2 neuraminidase. *Nat. Commun.* 13, 6443. <https://doi.org/10.1038/s41467-022-34060-8>.
14. Fowler, D.M., and Fields, S. (2014). Deep mutational scanning: a new style of protein science. *Nat. Methods* 11, 801–807. <https://doi.org/10.1038/nmeth.3027>.
15. Starr, T.N., Greaney, A.J., Hilton, S.K., Ellis, D., Crawford, K.H.D., Dingens, A.S., Navarro, M.J., Bowen, J.E., Tortorici, M.A., Walls, A.C., et al. (2020). Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* 182, 1295–1310.e20. <https://doi.org/10.1016/j.cell.2020.08.012>.
16. Greaney, A.J., Starr, T.N., Gilchuk, P., Zost, S.J., Binshtein, E., Loes, A.N., Hilton, S.K., Huddleston, J., Eguia, R., Crawford, K.H.D., et al. (2021). Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. *Cell Host Microbe* 29, 44–57.e9. <https://doi.org/10.1016/j.chom.2020.11.007>.

17. Greaney, A.J., Loes, A.N., Crawford, K.H.D., Starr, T.N., Malone, K.D., Chu, H.Y., and Bloom, J.D. (2021). Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe* 29, 463–476.e6. <https://doi.org/10.1016/j.chom.2021.02.003>.
18. Ellis, D., Brunette, N., Crawford, K.H.D., Walls, A.C., Pham, M.N., Chen, C., Herpoldt, K.L., Fiala, B., Murphy, M., Pettie, D., et al. (2021). Stabilization of the SARS-CoV-2 spike receptor-binding domain using deep mutational scanning and structure-based design. *Front. Immunol.* 12, 710263. <https://doi.org/10.3389/fimmu.2021.710263>.
19. Thyagarajan, B., and Bloom, J.D. (2014). The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. *Elife* 3, e03300. <https://doi.org/10.7554/eLife.03300>.
20. Doud, M.B., and Bloom, J.D. (2016). Accurate measurement of the effects of all amino-acid mutations on influenza hemagglutinin. *Viruses* 8, E155. <https://doi.org/10.3390/v8060155>.
21. Lee, J.M., Huddleston, J., Doud, M.B., Hooper, K.A., Wu, N.C., Bedford, T., and Bloom, J.D. (2018). Deep mutational scanning of hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants. *Proc. Natl. Acad. Sci. USA* 115, E8276–E8285. <https://doi.org/10.1073/pnas.1806133115>.
22. Wu, N.C., Young, A.P., Al-Mawsawi, L.Q., Olson, C.A., Feng, J., Qi, H., Chen, S.H., Lu, I.H., Lin, C.Y., Chin, R.G., et al. (2014). High-throughput profiling of influenza A virus hemagglutinin gene at single-nucleotide resolution. *Sci. Rep.* 4, 4942. <https://doi.org/10.1038/srep04942>.
23. Doud, M.B., Hensley, S.E., and Bloom, J.D. (2017). Complete mapping of viral escape from neutralizing antibodies. *PLoS Pathog.* 13, e1006271. <https://doi.org/10.1371/journal.ppat.1006271>.
24. Wu, N.C., Thompson, A.J., Lee, J.M., Su, W., Arlian, B.M., Xie, J., Lerner, R.A., Yen, H.L., Bloom, J.D., and Wilson, I.A. (2020). Different genetic barriers for resistance to HA stem antibodies in influenza H3 and H1 viruses. *Science* 368, 1335–1340. <https://doi.org/10.1126/science.aaz5143>.
25. Neumann, G., Watanabe, T., Ito, H., Watanabe, S., Goto, H., Gao, P., Hughes, M., Perez, D.R., Donis, R., Hoffmann, E., et al. (1999). Generation of influenza A viruses entirely from cloned cDNAs. *Proc. Natl. Acad. Sci. USA* 96, 9345–9350.
26. Tamura, D., Nguyen, H.T., Sleeman, K., Levine, M., Mishin, V.P., Yang, H., Guo, Z., Okomo-Adhiambo, M., Xu, X., Stevens, J., and Gubareva, L.V. (2013). Cell culture-selected substitutions in influenza A(H3N2) neuraminidase affect drug susceptibility assessment. *Antimicrob. Agents Chemother.* 57, 6141–6146. <https://doi.org/10.1128/AAC.01364-13>.
27. Hom, N., Gentles, L., Bloom, J.D., and Lee, K.K. (2019). Deep mutational scan of the highly conserved influenza A virus M1 matrix protein reveals substantial intrinsic mutational tolerance. *J. Virol.* 93, e00161-19. <https://doi.org/10.1128/JVI.00161-19>.
28. Soh, Y.S., Moncla, L.H., Eguia, R., Bedford, T., and Bloom, J.D. (2019). Comprehensive mapping of adaptation of the avian influenza polymerase protein PB2 to humans. *Elife* 8, e45079. <https://doi.org/10.7554/eLife.45079>.
29. Tokuriki, N., Stricher, F., Schymkowitz, J., Serrano, L., and Tawfik, D.S. (2007). The stability effects of protein mutations appear to be universally distributed. *J. Mol. Biol.* 369, 1318–1332. <https://doi.org/10.1016/j.jmb.2007.03.069>.
30. Pace, C.N., Shirley, B.A., McNutt, M., and Gajiwala, K. (1996). Forces contributing to the conformational stability of proteins. *FASEB J* 10, 75–83. <https://doi.org/10.1096/fasebj.10.1.8566551>.
31. Pakula, A.A., and Sauer, R.T. (1989). Genetic analysis of protein stability and function. *Annu. Rev. Genet.* 23, 289–310. <https://doi.org/10.1146/annurev.ge.23.120189.001445>.
32. Saito, T., Taylor, G., and Webster, R.G. (1995). Steps in maturation of influenza A virus neuraminidase. *J. Virol.* 69, 5011–5017. <https://doi.org/10.1128/JVI.69.8.5011-5017.1995>.
33. Wu, Z.L., Ethen, C., Hickey, G.E., and Jiang, W. (2009). Active 1918 pandemic flu viral neuraminidase has distinct N-glycan profile and is resistant to trypsin digestion. *Biochem. Biophys. Res. Commun.* 379, 749–753. <https://doi.org/10.1016/j.bbrc.2008.12.139>.
34. Delgado, J., Radusky, L.G., Cianferoni, D., and Serrano, L. (2019). FoldX 5.0: working with RNA, small molecules and a new graphical interface. *Bioinformatics* 35, 4168–4169. <https://doi.org/10.1093/bioinformatics/btz184>.
35. Radusky, L.G., and Serrano, L. (2022). pyFoldX: enabling biomolecular analysis and engineering along structural ensembles. *Bioinformatics* 38, 2353–2355. <https://doi.org/10.1093/bioinformatics/btac072>.
36. Hoie, M.H., Cagiada, M., Beck Frederiksen, A.H., Stein, A., and Lindorff-Larsen, K. (2022). Predicting and interpreting large-scale mutagenesis data using analyses of protein stability and conservation. *Cell Rep.* 38, 110207. <https://doi.org/10.1016/j.celrep.2021.110207>.
37. Echave, J., and Wilke, C.O. (2017). Biophysical models of protein evolution: understanding the patterns of evolutionary sequence divergence. *Annu. Rev. Biophys.* 46, 85–103. <https://doi.org/10.1146/annurev-biophys-070816-033819>.
38. Bastolla, U., Dehouck, Y., and Echave, J. (2017). What evolution tells us about protein physics, and protein physics tells us about evolution. *Curr. Opin. Struct. Biol.* 42, 59–66. <https://doi.org/10.1016/j.sbi.2016.10.020>.
39. Bloom, J.D., Labthavikul, S.T., Otey, C.R., and Arnold, F.H. (2006). Protein stability promotes evolvability. *Proc. Natl. Acad. Sci. USA* 103, 5869–5874. <https://doi.org/10.1073/pnas.0510098103>.
40. Rao, R., Liu, J., Verkuil, R., Meier, J., Canny, J.F., Abbeel, P., Sercu, T., and Rives, A. (2021). MSA transformer. Preprint at bioRxiv. <https://doi.org/10.1101/2021.02.12.430858>.
41. Wittmann, B.J., Yue, Y., and Arnold, F.H. (2021). Informed training set design enables efficient machine learning-assisted directed protein evolution. *Cell Syst.* 12, 1026–1045.e7. <https://doi.org/10.1016/j.cels.2021.07.008>.
42. Qiu, Y., Hu, J., and Wei, G.-W. (2021). Cluster learning-assisted directed evolution. *Nat. Comput. Sci.* 1, 809–818. <https://doi.org/10.1038/s43588-021-00168-y>.
43. Sanders, R.W., and Moore, J.P. (2021). Virus vaccines: proteins prefer prolines. *Cell Host Microbe* 29, 327–333. <https://doi.org/10.1016/j.chom.2021.02.002>.
44. Ellis, D., Lederhofer, J., Acton, O.J., Tsybovsky, Y., Kephart, S., Yap, C., Gillespie, R.A., Creanga, A., Olshefsky, A., Stephens, T., et al. (2022). Structure-based design of stabilized recombinant influenza neuraminidase tetramers. *Nat. Commun.* 13, 1825. <https://doi.org/10.1038/s41467-022-29416-z>.
45. Zhu, X., Turner, H.L., Lang, S., McBride, R., Bangaru, S., Gilchuk, I.M., Yu, W., Paulson, J.C., Crowe, J.E., Jr., Ward, A.B., and Wilson, I.A. (2019). Structural basis of protection against H7N9 influenza virus by human anti-N9 neuraminidase antibodies. *Cell Host Microbe* 26, 729–738.e4. <https://doi.org/10.1016/j.chom.2019.10.002>.
46. Kosik, I., and Yewdell, J.W. (2019). Influenza hemagglutinin and neuraminidase: yin-yang proteins coevolving to thwart immunity. *Viruses* 11, 346. <https://doi.org/10.3390/v11040346>.
47. Wagner, R., Matrosovich, M., and Klenk, H.D. (2002). Functional balance between haemagglutinin and neuraminidase in influenza virus infections. *Rev. Med. Virol.* 12, 159–166. <https://doi.org/10.1002/rmv.352>.
48. de Vries, E., Du, W., Guo, H., and de Haan, C.A.M. (2020). Influenza A virus hemagglutinin-neuraminidase-receptor balance: preserving virus motility. *Trends Microbiol.* 28, 57–67. <https://doi.org/10.1016/j.tim.2019.08.010>.
49. Liu, T., Wang, Y., Tan, T.J.C., Wu, N.C., and Brooke, C.B. (2022). The evolutionary potential of influenza A virus hemagglutinin is highly constrained by epistatic interactions with neuraminidase. *Cell Host Microbe* 30, 1363–1369.e4. <https://doi.org/10.1016/j.chom.2022.09.003>.
50. Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., Womack, T., and Bricogne, G. (2011). Data processing and analysis with

- the autoPROC toolbox. *Acta Crystallogr. D Biol. Crystallogr.* 67, 293–302. <https://doi.org/10.1107/S0907444911007773>.
51. Adams, P.D., Afonine, P.V., Bunkóczy, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* 66, 213–221. <https://doi.org/10.1107/S0907444909052925>.
 52. Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* 67, 355–367. <https://doi.org/10.1107/S0907444911001314>.
 53. Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of coot. *Acta Crystallogr. D Biol. Crystallogr.* 66, 486–501. <https://doi.org/10.1107/S0907444910007493>.
 54. Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 66, 12–21. <https://doi.org/10.1107/S0907444909042073>.
 55. Kinde, I., Wu, J., Papadopoulos, N., Kinzler, K.W., and Vogelstein, B. (2011). Detection and quantification of rare mutations with massively parallel sequencing. *Proc. Natl. Acad. Sci. USA* 108, 9530–9535. <https://doi.org/10.1073/pnas.1105422108>.
 56. Zhang, T.H., Wu, N.C., and Sun, R. (2016). A benchmark study on error-correction by read-pairing and tag-clustering in amplicon-based deep sequencing. *BMC Genom.* 17, 108. <https://doi.org/10.1186/s12864-016-2388-9>.
 57. Mölder, F., Jablonski, K.P., Letcher, B., Hall, M.B., Tomkins-Tinch, C.H., Sochat, V., Forster, J., Lee, S., Twardziok, S.O., Kanitz, A., et al. (2021). Sustainable data analysis with Snakemake. *F1000Res.* 10, 33. <https://doi.org/10.12688/f1000research.29032.2>.
 58. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *J.* 17, 10. <https://doi.org/10.14806/ej.17.1.200>.
 59. Magoč, T., and Salzberg, S.L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957–2963. <https://doi.org/10.1093/bioinformatics/btr507>.
 60. Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., and de Hoon, M.J.L. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25, 1422–1423. <https://doi.org/10.1093/bioinformatics/btp163>.
 61. Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577–2637. <https://doi.org/10.1002/bip.360221211>.
 62. Tien, M.Z., Meyer, A.G., Sydykova, D.K., Spielman, S.J., and Wilke, C.O. (2013). Maximum allowed solvent accessibilities of residues in proteins. *PLoS One* 8, e80635. <https://doi.org/10.1371/journal.pone.0080635>.
 63. Tareen, A., and Kinney, J.B. (2020). Logomaker: beautiful sequence logos in Python. *Bioinformatics* 36, 2272–2274. <https://doi.org/10.1093/bioinformatics/btz921>.
 64. Shu, Y., and McCauley, J. (2017). GISAIID: global initiative on sharing all influenza data - from vision to reality. *Euro Surveill.* 22, 30494. <https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494>.
 65. Meier, J., Rao, R., Verkuil, R., Liu, J., Seracu, T., and Rives, A. (2021). Language models enable zero-shot prediction of the effects of mutations on protein function. Preprint at bioRxiv. <https://doi.org/10.1101/2021.07.09.450648>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
6x-His Tag Monoclonal Antibody (HIS.H8)	Thermo Fisher Scientific	Cat#14-6657-82; RRID: AB_2572898
HRP Rat Anti-Mouse Ig, κ Light Chain	BD Biosciences	Cat#559751; RRID: AB_397315
Chemicals, peptides, and recombinant proteins		
Sodium chloride (NaCl)	Fisher Scientific	Cat#S271-500
Hydrochloric Acid, ACS, 12 M	Fisher Scientific	Cat#S25358
Tris Base	Fisher Scientific	Cat#BP152-500
Imidazole	Fisher Scientific	Cat#A10221
Calcium Chloride, Dihydrate	Millipore	Cat#208291-250GM
DpnI	New England Biolabs	Cat#R0176L
BsmBI-v2	New England Biolabs	Cat#R0739L
T4 DNA Ligase	New England Biolabs	Cat#M0202L
Lipofectamine 2000	Thermo Fisher Scientific	Cat#11668-019
Cellfectin II Reagent	Gibco	Cat#10362-100
TPCK-Trypsin	Thermo Fisher Scientific	Cat#20233
RNaseOUT	Thermo Fisher Scientific	Cat#10777019
6'-a-Sialyl-N-acetylglucosamine sodium salt	Biosynth	Cat#OS09313
JCSG Core Suite 1 Screen	Rigaku	Cat#1009842
JCSG Core Suite 2 Screen	Rigaku	Cat#1009843
JCSG Core Suite 3 Screen	Rigaku	Cat#1009844
JCSG Core Suite 4 Screen	Rigaku	Cat#1009845
Sf-900 II SFM	Thermo Fisher Scientific	Cat#10902088
DMEM medium	Thermo Fisher Scientific	Cat#11995065
Opti-MEM I Reduced Serum Medium	Thermo Fisher Scientific	Cat#31985070
GlutaMAX Supplement	Thermo Fisher Scientific	Cat#35050061
NEAA mixture (100x)	Lonza	Cat#13-114E
Trypsin-EDTA (0.25%), phenol red	Thermo Fisher Scientific	Cat#25200056
Penicillin-Streptomycin	Thermo Fisher Scientific	Cat#15140122
Fetal Bovine Serum (FBS)	Thermo Fisher Scientific	Cat#16000044
Phosphate-buffered saline (PBS), 1X	VWR	Cat#21-040-CM
NEB 5-alpha Competent <i>E. coli</i>	New England Biolabs	Cat#C2987H
MAX Efficiency DH10Bac Competent Cells	Thermo Fisher Scientific	Cat#10361012
MegaX DH10B T1R Electrocomp Cells	Thermo Fisher Scientific	Cat#C640003
NA protein sequences	GISAID; http://gisaid.org/	N/A
Ni Sepharose excel resin	Cytiva	Cat#17371202
XK 16/100 Superdex 200 pg	Cytiva	Cat#90100137
Critical commercial assays		
PrimeSTAR Max DNA Polymerase	Takara	Cat#R045A
KOD Hot Start DNA Polymerase	EMD Millipore	Cat#71086-3
PureLink PCR Purification Kit	Thermo Fisher Scientific	Cat#K310002
QIAprep Spin Miniprep Kit	Qiagen	Cat#27106
PureLink HiPure Plasmid Miniprep Kit	Thermo Fisher Scientific	Cat#K210003
Monarch DNA Gel Extraction Kit	New England Biolabs	Cat#T1020L
Superscript III reverse transcriptase	Thermo Fisher Scientific	Cat#18080044

(Continued on next page)

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Continued		
Deposited data		
Raw sequencing reads	This study	BioProject: PRJNA857746
X-ray coordinates and structure factors	This study	PDB: 8DWB
Experimental models: Cell lines		
Sf9 cells	ATCC	ATCC CRL-1711
MDCK-SIAT1 cells	Sigma-Aldrich	Cat#05071502-1VL
HEK 293T cells	N/A	N/A
Recombinant DNA		
pHW2000-chimeric HK68 HA	This study	N/A
pHW2000-Mos99 NA	This study	N/A
WSN 8-plasmid reverse genetics	Neumann et al. ²⁵	N/A
pFastBac-Mos99 NA	Lei et al. ¹³	N/A
Software and algorithms		
Octet analysis software 9.0	Fortebio	N/A
Python	https://www.python.org/	N/A
R	https://www.r-project.org/	N/A
autoPROC	Vonrhein et al. ⁵⁰	N/A
Phenix suite	Adams et al. ⁵¹	N/A
REFMAC5	Murshudov et al. ⁵²	N/A
COOT	Emsley et al. ⁵³	N/A
MolProbity	Chen et al. ⁵⁴	N/A
FoldX	Delgado et al. ³⁴	N/A
pyFoldX	Radusky and Serrano ³⁵	N/A
MSA Transformer	Rao et al. ⁴⁰	N/A
Custom scripts	This study	https://doi.org/10.5281/zenodo.7416749

RESOURCE AVAILABILITY

Lead contact

Information and requests for resources should be directed to and will be fulfilled by the lead contact, Nicholas C. Wu (nicwu@illinois.edu).

Materials availability

All plasmids generated in this study are available from the [lead contact](#) without restriction.

Data and code availability

- The read counts and fitness measurements from the deep mutational scanning experiment are in [Table S1](#). Raw sequencing data have been submitted to the NIH Short Read Archive under BioProject: [PRJNA857746](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA857746). The X-ray coordinates and structure factors have been deposited in the RCSB Protein DataBank under accession code PDB: 8DWB.
- Custom python scripts for analyzing the deep mutational scanning data have been deposited to <https://doi.org/10.5281/zenodo.7416749>.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODELS AND SUBJECT DETAILS

Cell cultures

HEK 293T cells (human embryonic kidney cells, female) were maintained in DMEM medium (Thermo Fisher Scientific) supplemented with 10% fetal bovine serum (FBS, Thermo Fisher Scientific), 1x MEM non-essential amino acids (Thermo Fisher Scientific), and 100 U mL⁻¹ of Penicillin-Streptomycin (Thermo Fisher Scientific). MDCK-SIAT1 cells (Madin-Darby canine kidney cells with stable expression of human 2,6-sialtransferase, female, Sigma-Aldrich) were maintained in DMEM medium supplemented with 10%

FBS, 1x MEM non-essential amino acids, and 100 U mL⁻¹ of Penicillin-Streptomycin. Sf9 cells (*Spodoptera frugiperda* ovarian cells, female, ATCC) were maintained in Sf-900 II SFM medium (Thermo Fisher Scientific).

METHOD DETAILS

Mutant library construction

The mutant library was constructed based on the pHW2000 eight-plasmid reverse genetics system for influenza virus.²⁵ Mos99 NA-encoding pHW2000 plasmid (pHW2000-Mos99-NA), which was cloned in our previous study,¹² was used as a PCR template to generate a linearized vector and a library of mutant Mos99 NA inserts. The linearized vector was generated using 5'-CGTACGTCTCAACACACGGAGCGCCCGGGGCCCTCT-3' and 5'-CGTACGTCTCATAACTGCTAGCGTTAACAGCTTGGG-3' as primers. Inserts were generated by two batches of PCRs, followed by overlapping PCRs. The first batch of PCRs consisted of 48 reactions, each with an equal molar mix of eight primers as the forward primer and a universal reverse primer 5'-CGTACGTCTCAGTGTGGCTGCATGGTGGCGTT-3'. The forward primers for the first batch of PCRs are listed in Table S4. These forward primers were named as cassetteX_N, in which X represents the cassette ID and N represents the primer number. Forward primers with the same cassette ID were mixed at equal molar ratio and used in the same PCR. The second batch of PCRs consisted of another 48 reactions, each with a universal forward primer 5'-CGTACGTCTCAGTTAACCGGAGTACTGGTCGAC-3' and a unique reverse primer as listed in Table S4. Subsequently, 48 overlapping PCRs were performed using the universal forward primer 5'-CGTACGTCTCAGTTAACCGGAGTACTGGTCGAC-3' and the universal reverse primer 5'-CGTACGTCTCAGTGTGGCTGCGATGGTGGCGTT-3'. For each overlapping PCR, the template was a mixture of 10 ng each of the corresponding products from the first and second batches of PCRs. The complete insert was an equal molar mix of the products of these 48 overlapping PCRs. All PCRs were performed using PrimeSTAR Max polymerase (Takara Bio) according to the manufacturer's instructions. PCR products were purified using Monarch DNA Gel Extraction Kit (New England Biolabs). Both the vector and the complete insert were digested by BsmBI (New England Biolabs) and ligated using T4 DNA ligase (New England Biolabs). The ligated product was transformed into MegaX DH10B T1R cells (Thermo Fisher Scientific). At least one million colonies were collected. Plasmid mutant libraries were purified from the bacteria colonies using PureLink HiPure Expi Plasmid Purification Kit (Thermo Fisher Scientific). The mutant library contained mutations from residues 82 to 465, which covered 99% of the NA head domain (residues 82 to 469).

Rescuing and passaging the mutant library

The virus mutant library was rescued by transfecting a co-culture of HEK 293T and MDCK-SIAT1 cells (ratio of 6:1) at 60% confluence in a T175 flask (175 cm²) with six internal segments (PB2, PB1, PA, NP, M, and NS) from H1N1 A/WSN/33, HA segment from H3N2 A/Hong Kong/1/1968, and the Mos99 NA mutant library. Transfection was performed using Lipofectamine 2000 (Thermo Fisher Scientific) according to the manufacturer's instructions. At 24 h post-transfection, cells were washed twice with PBS and cell culture medium was replaced with OPTI-MEM medium supplemented with 0.8 μg mL⁻¹ TPCK-trypsin. The virus mutant library was harvested at 72 h post-transfection, titered by TCID₅₀ assay using MDCK-SIAT1 cells, and stored at -80°C until used. To passage the virus mutant library, MDCK-SIAT1 cells in a T75 flask at 80% confluence were washed twice with PBS and infected with the virus mutant library at an MOI of 0.02 in OPTI-MEM medium supplemented with 0.8 μg mL⁻¹ TPCK-trypsin. At 2 h post-infection, infected cells were washed twice with PBS and fresh OPTI-MEM medium supplemented with 0.8 μg mL⁻¹ TPCK-trypsin was added to the cells. At 24 h post-infection, supernatant was harvested. Each replicate was transfected and passaged independently.

Sequencing library preparation

Viral RNA of the post-passaged mutant library was extracted from the supernatant using QIAamp Viral RNA Mini Kit (Qiagen). The extracted RNA was then reverse transcribed to cDNA using Superscript III reverse transcriptase (Thermo Fisher Scientific). The wild-type (WT) plasmid, plasmid mutant library, and the cDNA from the post-infection viral mutant library were amplified by PCR to add part of the adapter sequence required for Illumina sequencing. Six amplicon PCRs were performed for each sample using the following primers:

Amplicon 1: 5'-CACTCTTCCCTACACGACGCTCTCCGATCTNNNNNNNAAGGAAATATGCCCAAATA-3' and 5'-GACTGGA GTTCAGACGTGTGCTCTTCCGATCTNNNNNNNCTATCATGTACTGTGTCATT-3'

Amplicon 2: 5'-CACTCTTCCCTACACGACGCTCTCCGATCTNNNNNNNACACTAAACAACGGGCATTCA-3' and 5'-GACTGGA GTTCAGACGTGTGCTCTTCCGATCTNNNNNNNACCAATACTATCTACAAGCCT-3'

Amplicon 3: 5'-CACTCTTCCCTACACGACGCTCTCCGATCTNNNNNNNGCTAGCTTCATTTACAATGGG-3' and 5'-GACTGGA GTTCAGACGTGTGCTCTTCCGATCTNNNNNNNACAGGAGCATTCTCGACATG-3'

Amplicon 4: 5'-CACTCTTCCCTACACGACGCTCTCCGATCTNNNNNNNCCATTGTGTCAGGAAGTGCTCAG-3' and 5'-GACTGGA GTTCAGACGTGTGCTCTTCCGATCTNNNNNNNTTCTCATTGTAGGATCAA-3'

Amplicon 5: 5'-CACTCTTCCCTACACGACGCTCTCCGATCTNNNNNNNAGCTCCAGCAGTAGCCATTGC-3' and 5'-GACTGGA GTTCAGACGTGTGCTCTTCCGATCTNNNNNNNACCAGAATAACCGGACCTATT-3'

Amplicon 6: 5'-CACTCTTCCCTACACGACGCTCTCCGATCTNNNNNNNCAAGTCATAGTTGACAGAGGT-3' and 5'-ACTGGA GTTCAGACGTGTGCTCTTCCGATCTNNNNNNNGAACAAATTATAGGCATGAG-3'

A total of 12 N were included in each amplicon product as unique molecular identifiers (UMIs) to distinguish true mutations from sequencing errors.^{22,55,56} For each sample, the six amplicon PCR products were mixed at equal molar ratio. Subsequently, 3.2 million copies of mixed amplicon PCR products were used as template for a second PCR to add the rest of the adapter sequence and index to the amplicon using primers: 5'-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACX XXX XXX XAC ACT CTT TCC CTA CAC GAC GCT-3', and 5'-CAA GCA GAA GAC GGC ATA CGA GAT XXX XXX XXG TGA CTG GAG TTC AGA CGT GTG CT-3'. Positions annotated by an "X" represented the nucleotides for the index sequence. All PCRs were performed using KOD Hot Start DNA polymerase (MilliporeSigma) according to the manufacturer's instructions. PCR products were purified using PureLink PCR Purification Kit (Thermo Fisher Scientific). The final PCR products were submitted for the next-generation sequencing using one lane of NovaSeq SP PE150.

Sequencing data analysis

Sequencing data was obtained in FASTQ format and analyzed using a custom snakemake pipeline.⁵⁷ Firstly, paired-end reads with the same UMI were grouped together. Groups with only one read were discarded. Groups with reads that had different lengths were also discarded. A consensus read was generated for each group of reads with an 80% minimum frequency cutoff. In other words, for each position, the same nucleotide needed to be shared among at least 80% of the reads, otherwise the group would be discarded. Subsequently, primer sequences were trimmed using cutadapt⁵⁸ and each consensus read was assigned to one of the six amplicons based on the primer sequence. Consensus reads that were not assigned to any amplicon were discarded. Forward and reverse consensus reads were then merged by FLASH⁵⁹ using parameters: -m 30 -M 70 -l. Merged consensus reads were parsed by SeqIO module in BioPython⁶⁰ and translated into amino acid sequences. A merged consensus reads were removed if its sequence length was different from the referenced amplicon. Afterward, mutations were called by comparing the amino acid sequences of each merged consensus read to that of the referenced amplicon. Merged consensus reads that contained more than one amino acid mutations would be discarded. The frequency of each mutation in individual samples was computed. The frequency of mutation i in sample s was calculated as follows:

$$frequency_{i,s} = \frac{read\ count_{i,s} + 1}{\sum_{k \in s} read\ count_{k,s} + 1} \quad (1)$$

A mutation would be discarded if its frequency in the plasmid mutant library was not at least 6-fold higher than its frequency in the WT sample. This filter aimed to prevent error hotspots in reverse transcription, PCR, and next-generation sequencing from confounding our fitness calculation. In addition, a mutation would be discarded if its read count in the plasmid mutant library was 10 or less. After these filter steps, fitness calculation was performed.

For each replicate, the enrichment and fitness of a mutation i in amplicon amp was calculated as follows:

$$enrichment_i = \frac{read\ count_{post-passaged,i} + 1}{read\ count_{input,i} + 1} \quad (2)$$

$$fitness_i = \log_{10}(enrichment_i) - \frac{1}{N_{silent,amp} \sum_{k \in silent,amp} \log_{10}(enrichment_k)} \quad (3)$$

where the $read\ count_{input,i}$ represents the read count of mutation i in the plasmid mutant library, and $read\ count_{post-passaged,i}$ represents the read count of mutation i in the post-passaged mutant library. A pseudocount of 1 was added to the read counts to avoid division by zero. The fitness for mutation i was in \log_{10} scale and normalized to the mean fitness of silent mutations in amplicon amp , such that the mean fitness of silent mutations was 0. The final fitness was computed as the averaged fitness from the two biological replicates.

Protein expression and purification

The NA head domains, which contained residues 82 to 469, were fused to an N-terminal gp67 signal peptide, 6 × His-tag, a vasodilator-stimulated phosphoprotein (VASP) tetramerization domain, and a thrombin cleavage site. Recombinant bacmid DNA that carried Mos99 NA ectodomain was generated using the Bac-to-Bac system (Thermo Fisher Scientific) according to the manufacturer's instructions. Baculovirus was generated by transfecting the purified bacmid DNA into adherent Sf9 cells using Cellfectin reagent (Thermo Fisher Scientific) according to the manufacturer's instructions. The baculovirus was further amplified by passaging in adherent Sf9 cells at a multiplicity of infection (MOI) of 1. Recombinant NA head domains were expressed by infecting 1 L of suspension Sf9 cells at an MOI of 1. On day 3 post-infection, Sf9 cells were pelleted by centrifugation at 4,000 × g for 25 min, and soluble recombinant NA was purified from the supernatant by affinity chromatography using Ni Sepharose excel resin (Cytiva) and then size exclusion chromatography using a HiLoad 16/100 Superdex 200 prep grade column (Cytiva) in 20 mM Tris-HCl pH 8.0, 100 mM NaCl, and 10 mM CaCl₂. For crystallography, recombinant NA was further digested by thrombin (MilliporeSigma) for three weeks in 4°C using 15 U thrombin per mg of recombinant NA. The thrombin-digested recombinant NA was incubated with TALON metal affinity resin (Takara) for 2 h. The thrombin-digested recombinant NA in the flow-through and 10 mM imidazole wash was purified by size

exclusion chromatography using a HiLoad 16/100 Superdex 200 prep grade column (Cytiva) in 20 mM Tris pH 8.0, 100 mM NaCl, and 10 mM CaCl₂.

Crystallization and structural determination

Crystallization screening was performed using the JCSG Core Suites I-IV (Rigaku) with thrombin-digested NA at 7 mg mL⁻¹. Sitting drop for crystallization screening was set up by equal volume of precipitant and protein solution using the Crystal Gryphon (Art Robbins Instruments). Crystallization screens were incubated at 18°C. Initial hits were further optimized using the sitting drop method at 18°C, with 350 μL reservoir solution and 1:1 ratio of precipitant and protein solution. The crystallization condition was optimized to 1.9 M NH₄SO₄, 0.2 M Li₂SO₄, 0.1 M Tris pH 7.0. Large crystals grew over 3 days and were subsequently soaked for 3 h at 18°C in the precipitant solution supplemented with 20 mM N-acetylneuraminic acid. Crystals were soaked in precipitant solution supplemented with 20% glycerol as cryoprotectant prior to vitrification in liquid nitrogen.

Data were collected at the Advanced Photon Source (APS) at Argonne National Laboratory via the Life Science Collaborative Access Team (LS-CAT) at beamline 21-ID-D. Initial diffraction data were indexed, integrated, and scaled using autoPROC.⁵⁰ The structure was solved by molecular replacement using Phaser-MR included in the Phenix suite,⁵¹ with the apo Mos99 NA structure (PDB: 7U4F)¹³ as the replacement model. The structure was further refined using REFMAC5⁵² and was manually built in COOT.⁵³ Ramachandran statistics were calculated using MolProbity.⁵⁴

Residue classification

DSSP⁵¹ was used to compute the solvent exposure surface area (SASA) for individual residues in the Mos99 NA head domain (PDB: 7U4F).¹³ Relative solvent accessibility (RSA) was computed by dividing the SASA by the theoretical maximum allowed solvent accessibility of the corresponding amino acid.⁶² SASA and RSA were computed using both monomeric and tetrameric forms of Mos99 NA head domain. Active site was defined as previously described,⁶ which included residues 118, 151, 152, 224, 276, 292, 371, and 406. Similarly, antigenic regions were defined as previously described,¹⁰ which included residues 153, 197, 198, 199, 328, 329, 330, 331, 332, 333, 334, 335, 336, 339, 340, 341, 342, 343, 344, 345, 346, 347, 367, 368, 369, 370, 400, 401, 402, 403, 431, 432, 433, and 434. The remaining residues were classified into “buried”, “interface”, and “exposed”:

Buried residues: RSA_{monomer} < 0.2.

Interface residues: RSA_{monomer} ≥ 0.2 and (RSA_{monomer} - RSA_{tetramer})/RSA_{monomer} > 0.5.

Exposed: RSA_{monomer} ≥ 0.2 and (RSA_{monomer} - RSA_{tetramer})/RSA_{monomer} ≤ 0.5.

Sequence logo analysis

Sequence logos were generated by Logomaker in Python.⁶³ All sequences were downloaded from Global Initiative for Sharing Avian Influenza Data (GISAID).⁶⁴ Sequences with ambiguous amino acids were discarded. The representative strains for different NA subtypes were (GISAID accession numbers in parentheses):

- N1: H1N1 A/California/07/2009 (EPI185379)
- N2: H3N2 A/Moscow/10/1999 (EPI362897)
- N3: H2N3 A/swine/Missouri/2124514/2006 (EPI126570)
- N4: H10N4 A/mink/Sweden/E12665/84 (EPI16618)
- N5: H12N5 A/mallard-duck/ALB/60/1976 (EPI86400)
- N6: H3N6 A/chicken/Nanchang/7-010/2000 (EPI89011)
- N7: H10N7 A/mallard/ALB/196/1996 (EPI88485)
- N8: H3N8 A/duck/Ukraine/1/1963 (EPI3107)
- N9: H11N9 A/tern/Australia/G70C/1975 (EPI129556)

FoldX

We applied the FoldX force field³⁴ to predict the change of thermostability (ΔΔG) of NA protein upon single amino acid substitutions. The crystal structure of Mos99 NA in complex with sialic acid that was determined in this study was used as the input for FoldX. Atoms of sialic acid were parameterized using pyFoldX³⁵ with default parameters. We first prepared the complex structure by optimizing atomic details using FoldX RepairPDB. Next, we substituted each amino acid of the WT sequence to the other 19 amino acids with the FoldX BuildModel based on the optimized structure and output energy values before and after mutation. ΔΔG was computed by the difference of energy values between the WT and the mutant. The above procedure was repeated 10 times to obtain an average ΔΔG estimation. While our crystal structure represented the native tetrameric form of the Mos99 NA, we also performed the same ΔΔG prediction procedure for the monomeric form by using one protomer of the crystal structure as the input.

MSA transformer

MSA Transformer is a transformer protein language model pre-trained with the mask language modeling objective.⁴⁰ MSA Transformer takes a set of aligned sequences as input, and outputs the probability that an amino acid occurs at the masked position in a protein given the surrounding sequence context. It was trained on a dataset that contained 26 million multiple sequence alignments, each of which was generated by searching a Uniref. 50 sequence against the UniClust30 database with HHblits. The model was

trained on 32 V100 GPUs for 450,000 updates. To infer the mutational fitness of Mos99 NA, 66,562 full-length human H3N2 NA sequences from 1968 to 2020 from were downloaded from the GISAID database and used as the input MSA. Next, we masked individual positions in the Mos99 NA sequence one at a time and inferred the fitness for each single amino acid mutation at the given position by comparing its predicted probability to that of the WT amino acid⁶⁵:

$$\text{Inferred fitness} = \log p(x_i = x_i^{\text{mut}}) - \log p(x_i = x_i^{\text{WT}})$$

QUANTIFICATION AND STATISTICAL ANALYSIS

The p values reported in [Figure 2B](#) and all correlation coefficients in this study were computed by the paired Student's t-test using the R software package.

Update

Cell Reports

Volume 42, Issue 10, 31 October 2023, Page

DOI: <https://doi.org/10.1016/j.celrep.2023.113356>

Correction

Mutational fitness landscape of human influenza H3N2 neuraminidase

Ruipeng Lei, Andrea Hernandez Garcia, Timothy J.C. Tan, Qi Wen Teo, Yiquan Wang, Xiwen Zhang, Shitong Luo, Satish K. Nair, Jian Peng,* and Nicholas C. Wu*

*Correspondence: jianpeng@illinois.edu (J.P.), nicwu@illinois.edu (N.C.W.)
<https://doi.org/10.1016/j.celrep.2023.113356>

(Cell Reports 42, 111951; January 31, 2023)

In the originally published version of this article, the numerator and the denominator of Equation 2 were mistakenly swapped. The article has been updated to reflect the correct numerator and denominator, and the conclusions of the paper were unaffected by the initial error. The authors sincerely regret the error.

